Liver R2* is affected by both iron and fat: A dual biopsy-validated study of chronic liver disease

Markus Karlsson, Mattias Ekstedt, Nils Dahlström, Mikael Forsgren, Simone Ignatova, Bengt Norén, Olof Dahlqvist Leinhard, Stergios Kechagias and Peter Lundberg

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-158842

N.B.: When citing this work, cite the original publication.
https://doi.org/10.1002/jmri.26601

Original publication available at:
https://doi.org/10.1002/jmri.26601

Copyright: Wiley
http://eu.wiley.com/WileyCDA/
Liver R2* is Affected by Both Iron and Fat: A Dual Biopsy-Validated Study of Chronic Liver Disease

Abstract

Background
Liver iron content (LIC) in chronic liver disease (CLD) is currently determined by performing an invasive liver biopsy. MRI using R2* relaxometry is a non-invasive alternative for estimating LIC. Fat accumulation in the liver, or proton density fat fraction (PDFF), may be a possible confounder of R2* measurements. Previous studies of the effect of PDFF on R2*, have not used quantitative LIC measurement.

Purpose
To assess the associations between R2*, LIC, PDFF, and liver histology, in patients with suspected CLD.

Study Type
Prospective.

Population
81 patients with suspected CLD.

Field Strength/Sequence
1.5 T. Multi-echo turbo field echo to quantify R2*. PRESS MRS to quantify PDFF.

Assessment
Each patient underwent an MR-examination, followed by two needle biopsies immediately following the MR-examination. The first biopsy was used for conventional histological assessment of LIC, whereas the second biopsy was used to quantitatively measure LIC using mass spectrometry. R2* was correlated to both LIC and PDFF. A correction for the influence of fat on R2* was calculated.

Statistical Tests
Pearson correlation, linear regression, and area under the receiver operating curve.

Results
There was a positive linear correlation between R2* and PDFF (R = 0.69), after removing data from patients with iron overload, as defined by LIC. R2*, corrected for PDFF, was the best
method for identifying patients with elevated iron levels, with a correlation of $R = 0.87$ and a sensitivity and specificity of 87.5% and 98.6% respectively.

**Data Conclusion**

PDFF increases $R^*$.  

**Keywords**

Liver Iron Content, Liver Fat, $R^*$, PDFF, Iron Overload
Introduction

Iron overload is a potentially serious health problem associated with tissue damage in the liver, heart, joints, and endocrine organs.\(^1\) There are several causes for iron overload, such as multiple red blood cell transfusions, hereditary hemochromatosis, alcoholic liver disease and chronic hepatitis. Excess iron catalyzes the Haber-Weiss reaction leading to reactive oxygen species that are associated with inflammation and fibrosis.\(^2\) Symptoms of iron overload are generally mild and nonspecific and develop gradually.\(^3,4\) It is therefore important to identify iron overload before organ damage occurs.

Hyperferritinemia is a common finding during a medical assessment and is a common cause for referral to a hepatology department. Iron overload is associated with elevated serum ferritin levels, but elevated serum ferritin is not always associated with iron overload. In fact, in one study up to 70\% of patients referred to specialist centers because of elevated serum ferritin did not have iron overload.\(^5\) Common causes of non-iron overload hyperferritinemia include non-alcoholic fatty liver disease (NAFLD), alcohol consumption, and inflammatory disorders.\(^6\)

Traditionally, liver biopsy has been considered the reference standard for assessment of liver iron content (LIC). Hepatic iron can be determined either semi-quantitatively by a histopathologist using Perls' Prussian blue stain,\(^7\) or quantitatively using elemental analysis. Due to the invasive nature of biopsies, with potentially serious complications such as bleeding and perforation of adjacent organs, non-invasive measurements of hepatic iron are needed.

Several studies have used MRI to measure R2* relaxation rates as a non-invasive biomarker for LIC. Early studies mainly included patients receiving blood transfusions\(^8^{–}12\) and showed a linear relation between R2* and LIC. These patients typically have a very high LIC of 10-30 mg/g; in contrast, typical chronic liver disease (CLD) patients referred to a hepatology department are expected to have a much lower LIC of around or below 1.2 mg/g, which is considered the upper limit of normal LIC.\(^13\) Later work has studied R2* in other patient groups with lower levels of LIC, although few studies have compared R2* to both histology and quantitative elemental analysis of LIC.\(^14^{–}18\)

Recently, both Mamidipalli et al.\(^19\) and Bashir et al.\(^20\) reported that fat accumulation in the liver correlates with a slight increase in liver R2*. Although the exact mechanism is unknown, Mamidipalli et al. hypothesized that, similar to iron, hepatic fat droplets induce local susceptibility
effects, leading to an increased R2*. However, none of these studies reported actual LIC. Therefore, it is still possible that they only measured a general correlation between PDFF and LIC and not an intrinsic specific effect of fat on R2*.

The aim of this study was to assess the associations among R2*, LIC, PDFF, and liver histology in consecutively enrolled unselected patients scheduled for liver biopsy at a single center liver clinic.

**Subjects and Methods**

**Subjects**

This prospective study was performed between 2007 and 2014. Patients were consecutively recruited after being referred to the Department of Gastroenterology and Hepatology at Linköping University Hospital, Linköping, Sweden. Inclusion criteria were chronically (> 6 months) elevated serum levels of alanine aminotransferase (ALT; > 71 U/L for men and > 45 U/L for women) and/or aspartate aminotransferase (AST; > 45 U/L for men and > 36 U/L for women) and/or alkaline phosphatase (ALP; > 106 U/L for men and women). All patients who, on clinical indication, or as part of a clinical study, needed a liver biopsy for histopathological evaluation were asked to participate in the study. Exclusion criteria included contraindications for MRI (the presence of electronic devices such as pacemakers, implants with ferromagnetic properties, pregnancy, and claustrophobia) and liver biopsy (presence of primary or secondary coagulative disorder, prothrombin time > 1.5 times the international normalized ratio, platelet count <50×10⁹/L, hepatic malignancy, and clinical signs of decompensated cirrhosis). The study included an MRI examination as well as obtaining a second liver biopsy specimen. This study was part of a larger one investigating how multimodal MR techniques can replace liver biopsies. The original sample size was estimated based using ³¹P-MRS, as a biomarker for liver fibrosis. The study was approved by the Regional Ethics Committee at XX University (Reference No. XX) and all patients provided written informed consent.

**Histology**

Two biopsies were obtained from each patient using a percutaneous approach with a 1.6 mm BioPince™ needle (Argon Medical Devices, Plano, USA) with ultrasonographic guidance. The second biopsy was obtained using the same percutaneous approach in the same session, meaning that the two biopsies were obtained no more than a few centimeters apart, in the same liver
The first biopsy was sent to histopathological examination, while the second biopsy was obtained for elemental analysis, including iron.

A single histopathologist, with more than 10 years of experience in liver pathology (SI) and blinded to the results of the MR analysis, examined the first biopsy, performing a semi-quantitative evaluation of iron content, fibrosis stage, and steatosis. Iron content was graded on a scale from 0-4 using Perl's Prussian blue stain where the different grades are defined as 0: iron granules absent or barely discernible at × 400 magnification; 1: iron granules barely discernible at × 250 magnification and easily confirmed at × 400 magnification; 2: discrete iron granules resolved at × 100 magnification; 3: discrete iron granules resolved at × 25 magnification; and 4: masses visible at × 10 magnification, or with the naked eye. Fibrosis stage was scored on a scale from 0-4 where the different stages are defined as F0: no fibrosis; F1: portal or perisinusoidal fibrosis; F2: portal and perisinusoidal fibrosis; F3: bridging fibrosis; and F4: cirrhosis. Steatosis was graded on a scale from 0-3 corresponding to fat deposition in < 5%, 5-33%, 34-66%, and > 67% of the hepatocytes.

Liver iron content analysis

The second biopsy was immediately frozen at -80 °C. The frozen biopsies were later freeze-dried, weighed and sent to an external laboratory (ALS Scandinavia AB, Luleå, Sweden) for analysis of LIC, as mg iron per gram dry tissue (mg/g). At the laboratory, the dried samples were digested by adding 2.50 mL ‘super pure’ HNO₃ and 0.25 mL H₂O₂ followed by a 30 min treatment at 170 °C in a microwave oven. Samples were then diluted to 5.00 mL with MilliQ ultrapure water, for inductively coupled plasma sector field mass-spectrometry analysis (ICP-SFMS). We defined elevated iron levels as LIC > 1.2 mg/g.

MR-measurements

In all but one of the patients, the MR examination was performed the same day, 30-60 minutes before the biopsy procedure. The MR examination was performed after overnight fasting using a Philips Achieva 1.5 T MRI scanner (Philips Healthcare, Best, The Netherlands).

The R₂* relaxation rate of the water protons in the liver was quantified from images acquired using an axial 3D multi-echo turbo filed echo sequence with spectral pre-saturation with inversion recovery fat saturation (Figure 2A). The imaging used the following parameters: repetition time (TR) = 26.0 ms, echo time (TE) = 4.6/9.2/13.8/18.4/23.0 ms, flip angle (FA) = 20°, spatial resolution = 1.25x1.25x5.00 mm³, and field of view (FOV) = 320x290x80 mm³.
Liver proton density fat fraction (PDFF) was measured by MRS (PRESS sequence, TR = 1500 ms, TE = 35 ms, volume of interest (VOI) = 30x30x30 mm$^3$, NSA = 8), as described previously.$^{24}$ The amplitudes of the water and fat resonances were quantified using jMRUI and the AMARES algorithm.$^{25}$ The amplitudes of the water and fat resonances were corrected for differences in T1 and T2 relaxation times. Fat T1 and T2 relaxation times were assumed to be 236 ms and 58.8 ms respectively, while water T1 was assumed to be 663 ms. T2 of water was measured, using a single-slice axial turbo gradient spin echo sequence with spectral pre-saturation with inversion recovery fat saturation using the following imaging parameters: TR = 190 ms, TE = 6/12/18/24/30/36/42/48/54/60/66/72 ms, FA = 90°, spatial resolution = 0.98x0.98x10.00 mm$^3$, FOV = 375x294x10 mm$^3$. Furthermore, we also accounted for the lipid resonances hidden underneath the water resonance according to.$^{26}$ All the above-mentioned sequences were acquired during expiration breath-hold.

For calculation of both $R2^*$ and T2, a single region of interest (ROI) was placed in the same area of the liver as the spectroscopy VOI by a single radiologist, (ND), with more than ten years of experience in abdominal imaging and blinded to the results of both the histology and mass-spectrometry (Figure 2). The spectroscopy VOI was not actively placed in the same segment as the subsequently obtained liver biopsy. $R2^*$ and T2 were calculated by fitting a mono-exponential decay function to the mean signal intensities from the ROIs at the different echo times. The $R2^*$ and T2 analysis was performed using custom software written in Mathematica (Wolfram Research, Champaign, Illinois, USA).

**Statistical analysis**

Correlations with ordinal variables were assessed using Spearman’s correlation coefficient, and correlation with continuous variables was assessed using Pearson’s correlation coefficient. Receiver operating characteristic (ROC) curve analysis was used to evaluate diagnostic performance with respect to diagnosing hepatic iron overload. One-way analysis of variance was used to test for significant differences in LIC among different subgroups of the cohort. Linear regression was used to estimate relation between LIC, $R2^*$ and PDFF. A significance level of 0.05 was used in the statistical analysis.
Results

Subjects
A total of 94 patients underwent the MRI examination and liver biopsies. However, one patient decided to abort the MRI examination and twelve patients had only one biopsy (clinical), leading to a final cohort of 81 patients. A flow chart of included patients is presented in Figure 1 and the clinical characteristics of the cohort are shown in Table 1.

Liver iron content in the patient population
LIC was measured in all patients. Figure 3 shows LIC in various subgroups of the cohort. Figure 3A shows that LIC increases significantly with increasing histological Perls' score ($p < 0.01$). However, there is an overlap between the patients with Perls' score = 0 and patients with Perls' score $\geq 1$. Fifty percent of patients staged with elevated iron levels by the pathologist had normal LIC levels, whereas one patient with Perls' score = 0 actually had LIC $> 1.2$ mg/g. Furthermore, there was also a significant difference ($p < 0.05$) in LIC between patients with elevated serum ferritin ($>300$ µg/L), and patients with normal serum ferritin (Figure 3C). As with Perls' score, there was also an overlap with respect to serum ferritin, as 68% of the patients with elevated serum ferritin did not have elevated iron levels as determined by LIC, whereas 3% of the patients with normal serum ferritin had elevated iron levels, as indicated by LIC. Moreover, there was no significant difference in LIC between the different patient etiologies (Figure 3B), nor were there any significant differences in LIC between patients with different fibrosis stages (Figure 3D).

Correlation between MRI and Liver Iron Content
There was a clear linear relation between $R2^*$ and LIC, except at the low end of the scale (Figure 4A; $R = 0.82$). A linear regression gave a slope of 0.027 (CI: 0.023 – 0.032) and an offset of -0.352 (CI: -0.521 – -0.183). If the outlying patient with hereditary hemochromatosis was removed from the analysis, the correlation was still moderate ($R = 0.68$) and a linear regression showed no significant differences in either slope, 0.026 (CI: 0.020 – 0.033), or offset -0.320 (CI: -0.555 – -0.086). A Bland-Altman analysis (Figure 4B) of the measured LIC and LIC estimated using the measured $R2^*$ values showed a 95% limit of agreement of ±0.60 mg/g, with the plot showing a larger variability for the lowest range of LIC. Finally, Figure 4C shows our calibration curve for $R2^*$ and LIC, compared to calibration curves obtained in other studies$^9$-$^11$, which mainly included blood transfusion patients with an entirely different range of LIC.
Associations between R2* and PDFF

We investigated whether R2* was associated to fat accumulation, which has previously been reported.\textsuperscript{19,20} Figure 5 shows the correlation between R2* and PDFF, measured by MRS. Since previous work has suggested that the effect of fat on R2* is smaller than the effect of iron\textsuperscript{19}, we excluded all patients with elevated iron levels (LIC > 1.2 mg/g). The result was a moderate, but significant correlation between R2* and PDFF (Figure 5A; R = 0.69) giving the following regression:

\[ R2^* = 28.6 + 1.04 \cdot PDFF, \]  
\[ [1] \]

indicating that for every percentage point of fat a patient has in the liver, R2* increases by close to one s\textsuperscript{-1}. To verify that the correlation between R2* and PDFF is not due to a correlation between fat and iron, we also calculated the correlation between PDFF and LIC, which was non-existent (R = 0.07).

A linear regression analysis with R2* as the response variable, and LIC and PDFF as predictors gave the following model:

\[ R2^* = 16.1 + 25.4 \cdot LIC + 0.96 \cdot PDFF, \]  
\[ [2] \]

with a confidence interval of the LIC-coefficient of 22.1 – 28.1 and a confidence interval of the PDFF-coefficient of 0.62 – 1.29. Again, this indicates that R2* increases by close to one s\textsuperscript{-1} for every percentage point of PDFF. This model was used to create a PDFF-corrected R2* (PDFF-R2*), by subtracting 0.96 times PDFF for each patient:

\[ PDFF-R2^* = R2^* - 0.96 \cdot PDFF. \]  
\[ [3] \]

When comparing the PDFF adjusted R2* to LIC, the correlation increased from R = 0.82 to R = 0.87 (Figure 5B), with a slope of 0.030 (CI: 0.026 – 0.034) and an offset of -0.325 (CI: -0.461 – -0.189). If the outlying patient was removed, the correlation increased from R = 0.68 to R = 0.74.

Comparison of methods for identifying elevated iron levels

Figure 6 shows an ROC-analysis, conducted to evaluate the ability of R2*, PDFF-R2*, and serum ferritin to identify patients with elevated iron levels. The figure shows that R2*, corrected for fat, is the best method.
Table 2 show sensitivities, specificities as well as positive and negative predictive values for Perls’ score, R2*, PDFF-R2*, and serum ferritin, for identifying patients with iron overload. As in Figure 6, the table indicates that PDFF-R2* is the best method. When using PDFF-R2* instead of R2*, two instead of five patients are misclassified.

**Discussion**

We have demonstrated a direct correlation between the R2* relaxation rate and PDFF in the liver, in a diverse cohort of CLD patients. Furthermore, by measuring actual LIC in biopsies, we have also verified that the correlation between R2* and PDFF is intrinsic and not just an effect of a correlation between PDFF and LIC.

Our results should be seen in the context of the recent findings by Mamidipalli et al.\(^{19}\), who investigated the correlation between R2* and PDFF in a cohort of 184 pediatric NAFLD patients. As we have here, they also found a similar correlation between R2* and PDFF. As their cohort was larger and only contained NAFLD patients, they have more patients covering the whole pathological range of PDFF. However, a limitation of their findings is that they could not obtain biopsies from their patients. Therefore, they could not measure LIC and thus not ascertain that the correlation between R2* and PDFF was not due to a correlation between PDFF and LIC. Herein, we did obtain biopsies. After removing all patients with elevated LIC, according to mass spectrometry, we could still see a correlation between R2* and PDFF, just as Mamidipalli and colleagues did. All in all, our combined findings strongly suggest that there is an intrinsic effect of PDFF on R2*.

What is the likely biophysical background to the observed effects of PDFF on R2* (or R2) measurements? The intracellular environment of hepatocytes contains a range of particles that will affect the relaxation of water in different ways. Large aggregates, particles, or spheres with a volume susceptibility that is typically lower, or higher, than the cytosol will create inhomogeneous broadening of the water resonance. Thus, they will create local, but macroscopic, static variability of the magnetic field across the sample, and will therefore only affect R2*. In contrast, compounds that are more homogenously distributed and motionally contributing to a fluctuating magnetic field, will result in homogenous broadening of the water resonance; in other words, they will affect R2.

The mechanistic background to the inhomogeneous broadening is illustrated schematically in Figure 7. Lipid spheres may have a size of c. 30 µm (Figure 7), whereas ferritin is typically small octahedral particles (c. 0.01 µm), which at least at relatively low LIC-levels, are evenly
distributed in the cytosol as visualized by Perls' staining. The latter objects may therefore contribute both to R2 and R2*, whereas large lipids spheres will only contribute to R2*, since the inhomogeneous broadening will be removed in R2-measurements, but not in R2*-measurements. Depending on the specific ratio between intra- versus extra-sphere volume susceptibilities, the affected area extends in the order of one sphere radius from the edge of, e.g., the lipid sphere. Hemosiderin is typically much larger than ferritin, about 1 µm, very heterogenous, and will likely therefore mainly contribute to R2*. A discussion of the differential relaxation effects from ferritin and hemosiderin in relation to R2 and R2* is available in.

We have chosen to measure R2* using fat-suppressed gradient echo images acquired at in-phase echo times, whereas Mamidipalli et al. used a chemical shift-based imaging (CSI) approach (CSI). A possible downside of our approach is that the fat suppression is imperfect and that our echo times could be too long to study patients with very high levels of LIC. Thalassemia patients, e.g., can have ten times more iron than we found in our cohort. If the LIC levels, and thus R2*, are too high, there might not be sufficient SNR to accurately estimate R2*. On the other hand, a particular advantage of our approach is that since we did not have to model the signals for fat and water, our results were not affected by a choice of fat-water signal model. All in all, the fact that both we and Mamidipalli et al. have been able to show the same effect using different imaging approaches confirms that the observed effect is real.

It can also be said that if an MR-based iron quantification is to be used in clinical practice, CSI would probably be preferable. The reasons for this are threefold. First, using CSI, R2* and PDFF maps can be acquired simultaneously, with full liver coverage in a single breath hold. Second, CSI is probably more widely available than spectroscopy and the postprocessing CSI-images is typically fully automated, while spectroscopy often require some manual processing. Last, CSI-sequences typically have shorter echo times, compared to our sequence. CSI should therefore be more accurate when examining patients with more severe iron overload, as mentioned above.

Our results suggest that R2* increases by around 1 s⁻¹ for every percentage point of PDFF at 1.5 T. Moreover, patients with severe steatosis can have a PDFF of up to 30%; R2* may thus be confounded by up to 30 s⁻¹. Since patients with severe iron overload can have a liver R2* of several hundred s⁻¹, considering the effect of PDFF may not be necessary, in such patients. On the other hand, in a cohort such as the one investigated here (where most patients had a liver R2* below 60 s⁻¹), considering the effect of PDFF would be more important, since a slight increase in R2* could be due to either a slight increase in iron, or steatosis. As long as a CSI...
approach is used for quantifying both R2* and PDFF, no further image acquisition will be required. However, these findings should be investigated further in cohorts including more patients with elevated iron levels.

We have shown that R2* can replace histology for evaluating LIC in patients with CLD. Not only is an MRI examination non-invasive, R2* also shows better sensitivity and specificity than histology. This could be in part because both macro- and microscopically heterogeneous distributions of iron exist in the liver tissue, but also because of the inherent difficulties of performing a semi-quantitative assessment of a biopsy specimen. These results are strengthened by the agreement we observed between our calibration curve and those found in studies of patients receiving blood cell transfusions\(^9\)-\(^{11}\) which suggests that our results are generalizable.

Only ten percent of our patients had elevated iron levels, even though we used a relatively low but reasonable cut-off of 1.2 mg/g\(^{13}\) While the exact limits for what is considered iron overload would have consequences for our exact numbers for specificity and sensitivity, this does not change our conclusion that R2* correlates linearly with LIC and that it can replace histology.

The composition of our cohort is both a strength and a limitation of our study. It is a strength in that our cohort is representative of non-selected CLD patients referred to a typical hepatology department. It is obvious that in a non-selected patient population scheduled for liver biopsy, few patients will have elevated iron levels. On the other hand, the low number of patients with iron overload is a weakness of our study in that it limits our power to statistically determine if the different methods are significantly different.

Another limitation of the study is that we used LIC measurements in biopsies as the gold standard for iron overload. Although LIC is much more accurate than semi-quantitative histology, it still suffers from the same sampling variability. A final limitation is that we did not obtain multiple histologic readings of the biopsies, to quantify inter-reader variability.

In conclusion, we have shown that liver R2* is affected by liver fat, and that fat increases R2*.
References

### Table 1. Demographic, clinical, and histopathological characteristics of the study cohort.

<table>
<thead>
<tr>
<th>All Patients (n = 81)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (n(%))</td>
</tr>
<tr>
<td>Age (Years)</td>
</tr>
<tr>
<td>Height (m)</td>
</tr>
<tr>
<td>Weight (kg)</td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
</tr>
<tr>
<td>Diabetes</td>
</tr>
<tr>
<td>AST (U/L)</td>
</tr>
<tr>
<td>ALT (U/L)</td>
</tr>
<tr>
<td>ALP (U/L)</td>
</tr>
<tr>
<td>Ferritin (µg/L)</td>
</tr>
<tr>
<td>Transferrin (g/L)</td>
</tr>
<tr>
<td>Serum Iron (µmol/L)</td>
</tr>
<tr>
<td>Transferrin Saturation (%)</td>
</tr>
<tr>
<td>Histological Steatosis Grade</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>Fibrosis Stage</td>
</tr>
<tr>
<td>F0</td>
</tr>
<tr>
<td>F1</td>
</tr>
<tr>
<td>F2</td>
</tr>
<tr>
<td>F3</td>
</tr>
<tr>
<td>F4</td>
</tr>
<tr>
<td>Histological Iron Grade (Perls)</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>Liver Iron Content (mg/g)</td>
</tr>
<tr>
<td>Iron Overload*</td>
</tr>
<tr>
<td>Diagnosis</td>
</tr>
<tr>
<td>Normal</td>
</tr>
<tr>
<td>NAFLD</td>
</tr>
<tr>
<td>HCV</td>
</tr>
<tr>
<td>PSC</td>
</tr>
<tr>
<td>PBC</td>
</tr>
<tr>
<td>AIH</td>
</tr>
<tr>
<td>Overlap AIH-PSC or AIH-PBC</td>
</tr>
<tr>
<td>Hemochromatosis</td>
</tr>
<tr>
<td>DILI</td>
</tr>
<tr>
<td>Wilson’s disease</td>
</tr>
<tr>
<td>ALD</td>
</tr>
<tr>
<td>AAT deficiency</td>
</tr>
</tbody>
</table>

*Liver Iron Content > 1.2 mg/g; Abbreviations: BMI: body mass index; AST: aspartate aminotransferase; ALT: alanine aminotransferase; ALP: alkaline phosphatase; NAFLD: non-alcoholic fatty liver disease; HCV: viral hepatitis-C; PSC: primary sclerosing cholangitis; PBC: primary biliary cholangitis; AIH: autoimmune hepatitis; DILI: drug-induced liver injury; ALD: alcoholic liver disease; AAT: alpha-1 antitrypsin
Table 2. Diagnostic performance of Perls' score, MRI and serum ferritin, with respect to identifying patients elevated iron levels (LIC>1.2 mg/g)

<table>
<thead>
<tr>
<th>Method</th>
<th>Cut-Off (s¹ or µg/L)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perls</td>
<td>1</td>
<td>87.5</td>
<td>90.4</td>
<td>50.0</td>
<td>98.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>100.0</td>
<td>90.4</td>
<td>53.3</td>
<td>100.0</td>
<td>0.969</td>
</tr>
<tr>
<td>R2*</td>
<td>48</td>
<td>87.5</td>
<td>94.5</td>
<td>63.6</td>
<td>98.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>25.0</td>
<td>100.0</td>
<td>100.0</td>
<td>92.4</td>
<td></td>
</tr>
<tr>
<td>PDFF-R2*</td>
<td>42</td>
<td>100.0</td>
<td>97.3</td>
<td>80.0</td>
<td>100.0</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>87.5</td>
<td>98.6</td>
<td>87.5</td>
<td>98.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>62.0</td>
<td>100.0</td>
<td>100.0</td>
<td>96.1</td>
<td></td>
</tr>
<tr>
<td>Ferritin</td>
<td>260</td>
<td>100.0</td>
<td>78.1</td>
<td>33.3</td>
<td>100.0</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>431</td>
<td>75.0</td>
<td>94.5</td>
<td>60.0</td>
<td>97.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>670</td>
<td>50.0</td>
<td>100.0</td>
<td>100.0</td>
<td>94.8</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: PPV: Positive predictive value; NPV: Negative predicted value; AUROC: Area under receiver operating characteristic curve
Figures

Figure 1. Study design.

Figure 2. Examples of the R2* quantification images at all echo times for one patient, as well as the R2* parameter fitting. The region of interest is shown in green, the MRS volume of interest is shown as the shaded area in the first image.
Figure 3. Liver iron content (LIC) in various patient subgroups. (A) LIC grouped by histological Perls’ score. (B) LIC grouped by disease etiologies. (C) LIC grouped by elevated and not elevated serum ferritin. (D) LIC grouped by fibrosis stage. The bars indicate medians and interquartile distances. The dashed line indicates a threshold for elevated iron levels (1.2 mg/g); *significant difference (p<0.05); NAFLD: non-alcoholic fatty liver disease; ALD: alcoholic liver disease; PSC: primary sclerosing cholangitis; PBC: primary biliary cirrhosis; AIH: autoimmune hepatitis.

Figure 4. Association between LIC and R2*. (A) Correlation between LIC and R2*. The different markers indicate the Perls’ score given by the histopathologist. The black line is a linear regression. (B) Bland-Altman analysis of LIC measured from the biopsy and LIC estimated using MRI, indicating that the error when estimating LIC using R2* is around 0.6 mg/g. (C) Our calibration curve for R2* and LIC compared to other calibration curves found in the literature. Note that the curve by Hankins is almost completely covered by the curve by Karlsson.
Figure 5. Relation between R2* and PDFF. (A) Correlation between R2* relaxation rate and PDFF. The black line is a linear regression. (B) Correlation between LIC and PDFF-corrected R2* (PDFF-R2*). The different markers indicate the Perls’ score given by the histopathologist. The black line is a linear regression.

Figure 6. Receiver operating characteristic (ROC) curves showing the ability of serum ferritin, R2*, and PDFF-corrected R2* (PDFF-R2*) to identifying patients with elevated iron levels. The curves indicate that PDFF-R2* is the best method.
Figure 7. Inhomogeneous broadening (contributing to R2*) by various intracellular structural components in hepatocytes. (A and B) Biopsy of a patient with a PDFF of c. 25%. Hematoxylin and Eosin-staining (H&E); grid cell length is 35 µm. (C-F) Schematic models of inhomogeneous broadening caused by differences in volume susceptibility between the interior of the particle, modeled as spheres, and the cytosolic environment surrounding the sphere. (C) ratio between volume susceptibilities inside vs. outside of the particle (\(\varepsilon\)) is 0.8 ('Lipid sphere'); (D) \(\varepsilon = 0.2\) ('Lipid sphere'); (E) \(\varepsilon = 10.0\) ('ferritin sphere'); (F) \(\varepsilon = 10.0\) ('hemosiderin sphere'). (The field lines were calculated using a Mathematica-script kindly provided by professor P. Kuchel, Sydney University, based on equations found in Lorrain and Corson, 'Electromagnetic fields and waves'.)